A Case Study of the Reasoning in a Genetics Experiment

Heuristic Programming Project
Working Paper 77-18

Jerry Feitelson
Mark Stefik

Abstract: The laboratory steps for a series of genetics
experiments are examined in depth and the reasoning and knowledge used
to plan the experiments are characterized. One surprise is the extent
to which the planning process seems to be event driven. For this
experiment, the planning process would not be well characterized as the
search of a large space for the solution to a fixed experiment.
Rather, most planning in these experiments seems to be short term and
in response to unexpected results in the laboratory. Considerable
knowledge is used in forming new hypotheses in response to the
unexpected. Furthermore, much of the geneticist's behavior seems to be
directed toward exploiting serendipity.

Table of Contents

Chapter I

Introduction

The MOLGEN [1] group has described a process of experiment planning which made much use of hierarchical problem solving to pre-plan the experimental steps. These experiments have been classified as synthesis or analysis experiments with goals explicitly defined by the user at the beginning of planning. This report explores in depth a series of actual genetics experiments [2] performed in Professor Joshua Lederberg's laboratory over a two month period in order to characterize both the experiment design process and the knowledge used to guide the decision making.

In this study, the geneticist is viewed as having a theory in which various beliefs are held with different degrees of certainty. His theory describes the limitations of his laboratory and his current research goals. Given this theory, he must decide what to do next. He may plan to do certain experiments and perform certain operations, but unexpected measurements and fortunate observations can steer him in different directions.

Judging only from this single case study, relatively little time is spent evaluating experimental alternatives for a fixed experiment. The art of successful experimentation involves shrewd decisions of what to do next based on recent observations and the state of the theory. A geneticist is an expert at exploiting serendipity and at generating hypotheses to be tested.

Chapter II gives some of the genetics background and terminology for the experiment being studied. Chapter III then elaborates the steps of the experiment in considerable detail. Chapter IV steps through the experiment again and highlights the knowledge that is used to guide decisions at critical moments of planning. Finally, Chapter V makes some general observations about the knowledge used and the planning process.

----

[1] See [Martin77] and [Stefik77].

[2] See [Ehrlich76] and [Ehrlich77].

Chapter II

Background and Some Definitions

The experiment described involves two kinds of bacteria -- Escherichia coli (E. coli) and Bacillus subtilis (B. subtilis) and the bacteriophage Phi-3-T. Phi-3-T carries a thymidylate synthetase gene. When Phi-3-T infects Thy⁻ B. subtilis (i.e. lacking in the ability to synthesize Thymidine), it "transforms it to prototrophy". This means that the Phi-3-T confers the ability to synthesize thymidine to the B. subtilis by donating its own genetic information. Thy⁻ B. subtilis requires thymidine in the growth medium to survive. We will be concerned with two plasmids (small, extrachromosomal, circular DNA molecules) of E. coli in addition to the main bacterial chromosome. These are called pSC101 and pMB9. Both plasmids carry a gene for resistance to the drug tetracycline. When E. coli lacks this gene, it is sensitive to tetracycline and is said to have the Tc^s phenotype. The plasmids have the ability to confer resistance to tetracycline when they are taken into E. coli -- giving it the Tc^r phenotype. They are used as "vectors", that is, carriers for foreign DNA.

B. subtilis	The bacterium Bacillus subtilis.
E. coli	The bacterium Escherichia coli.
Thy^+, Thy^-	Thymidine (T) is a base essential to synthesize or repair DNA. Thy⁻ bacteria require it in the growth medium. Thy⁺ bacteria can synthesize it from other compounds in the environment.
Tc^r, Tc^s	Resistance (or sensitivity) to tetracycline.
Phi-3-T	Bacteriophage which can lysogenically convert Thy⁻ B. subtilis to Thy⁺.
pSC101	Plasmid of E. coli carrying Tc^r.
pMB9	Another E. coli plasmid also carrying Tc^r.
pFT	Hybrid plasmid carrying Phi-3-T gene. Such plasmids will hybridize with Phi-3-T cRNA.
ColE1-amp	Another plasmid conferring resistance to the antibiotic ampicillin.
cRNA	Radioactively-labeled complementary RNA made from DNA templates using RNA polymerase.
$EcoR_I$	A restriction enzyme from E. coli.
$BamH_I$	A restriction enzyme from B. amyloliquefaciens.
T4 ligase	A ligase enzyme, which can join together DNA segments which have been cleaved by $EcoR_I$.
EM	Abbreviation for electron microscopy. This technique allows viewing many gross features of appropriately stained DNA molecules.
r^-m^-	The phenotype of the E. coli used in this

	experiment lacks a restriction/modification system for destroying foreign DNA. This makes the introduction of plasmids easier technically.
promoter	A region of DNA to which RNA polymerase binds to initiate transcription.
prototroph	Bacteria with the ability to grow on a minimal medium. ("Wildtype")
auxotroph	Bacteria unable to grow without nutritional supplements. This is often due to genetic defects at defined loci.

Chapter III

A Description of the Experiment

For the last two years, Prof. Lederberg's group has been trying to transfer a gene from B. subtilis to E. coli. In this experiment, they modified their goal somewhat by trying instead to transfer a gene from Phi-3-T, a bacteriophage of B. subtilis. DNA from the bacteriophage is somewhat easier to handle. It is shorter and easier to obtain in concentrated and purified amounts. The Thy gene in Phi-3-T is capable of restoring to prototrophy strains of B. subtilis which are deficient in either of the thymidine genes -- Thy A or Thy B. This experiment is significant because it supports the hypothesis that genes can be transferred between prokaryotes and expressed (ie. produce functional products in the new host). In addition, several interesting observations made during this experiment have suggested some directions for further research.

This chapter describes the experimental steps in detail but gives only scant motivation for the steps or interpretation of results. Chapter IV goes through this experiment again emphasizing the knowledge used in planning the steps and in analyzing the results.

III.1   Overview of the Experiment

I.   Extract Phi-3-T DNA from the bacteriophage (and perform various tests on it).

II. Digest Phi-3-T DNA with the enzyme $EcoR_I$ and test for transforming activity.

III. Ligate the Phi-3-T DNA with $EcoR_I$ cleaved pSC101 plasmids to create hybrid plasmids.

IV. Transform r^- m^- Thy^- Tc^S E. coli to Thy^+ Tc^r with the cloned plasmids and Isolate the transformants. (This involves reasoning based on biological function of the genes involved.)

V.   Verify that the Thy^+ character of the transformants is actually conferred by the hybrid plasmid. (Other hypotheses are possible and must be tested.)

VI. Use heteroduplex analysis to examine the molecular structure of

the hybrid plasmids to determine which segment contains the Thy gene.

VII.   Transfer the hybrid plasmids into B. subtilis and test for transformation.

VIII.   Since the technique of Heteroduplex Analysis is so time-consuming and not optimal for large bacterial chromosomes, a simple (albeit less specific) method for testing for the presence of DNA sequences by homology was extended. In particular, the in situ colony hybridization technique was tested and verified to work for B. subtilis.

IX.   Use the extended colony hybridization technique to examine the molecular structure of the transformed B. subtilis. A surprising result was that the Thy gene was incorporated into the B. subtilis chromosome but that pSC101 DNA was not.

X.   Reverse the hybridization procedure and to check if the hybrid plasmids are homologous to a single band in the Phi-3-T DNA.

The first four steps constitute the synthesis part of the experiment. These steps are designed to create colonies of E. coli with a foreign gene expressed. Steps V and VI are designed to test the synthesis steps and could be termed the analysis steps of the experiment. The final steps continue the analysis and explore some related matters. The next section reviews each of these steps in greater detail.

III.2   The Experimental Steps in Detail

I.   Extract and purify the Phi-3-T DNA.

   A.   Verify purity (satisfactory level of protein contamination) by measuring ratio of UV absorption.

   B.   Verify a satisfactory degree of intact molecules and molecular weight using EM.

   C.   Measure transforming ability in B. subtilis. (Adequate transforming activity is necessary to insure a reasonable expectation of success in later steps.)

II.   Digest Phi-3-T DNA with the enzyme $EcoR_I$ and test for transforming activity.

5

A.   Perform complete digestion with EcoR$_I$. (The motivation here
     is to get Thy gene on a short segment since the   complete phage
     is too long for introduction into the plasmid.   The resulting
     sticky ends will be useful in the later ligation step.)

B.   Use electrophoresis to obtain a cleavage pattern of the
      digestion products and to estimate the molecular weight.

     1.   Compare the digestion products to published cleavage
           pattern.

     2.   Compare the electrophoresis and EM estimates for molecular
           weight. Discrepancy noted in molecular weight measurement.
           (EM measurement is higher than electrophoresis measurement
           -- 72 million vs. 83 million.)

C.   Generate and test hypotheses for molecular weight discrepancy.
      (See Chapter IV for a discussion of the reasoning behind
      these hypotheses.) The following hypotheses were formed:

     1.   Incomplete gel resolution.

     2.   Loss of small electrophoresis fragments migrating out of
           gel.

     3.   Use of different standards for EM and gel-electrophoresis.
           For both techniques, molecular standards need to be used
           which will be distinguishable from what is being measured.

     4.   Repetition in the phage genome.   (This hypothesis was
           suggested in the paper, but is hard to understand.)

D.   Transforming activity of the Phi-3-T DNA was checked in B.
      subtilis. It was observed that the transforming activity was
      reduced by a factor of 1000 after complete EcoR$_I$ digestion.

E.   Generate hypotheses to explain the loss of transforming
      activity and test as follows:

     1.   Hypothesize that EcoR$_I$ damaged the Thy gene.   This
           hypothesis is disconfirmed by the fact that transforming
           activity continues after complete digestion.   Complete
           digestion was assumed by noticing no change   in the
           restriction pattern after increasing the digestion time by
           a factor of 10, and increasing the enzyme concentration by
           a factor of 10.

     2.   Hypothesize edge effects.   (This idea was discussed   in a
           thesis by Ron Harris but the effect has not yet been
           completely established.) This hypothesis was not pursued.

3.   Hypothesize that transforming activity decreases if the DNA pieces are too small.

a.   Decide to check the kinetics of $EcoR_I$ Phi-3-T DNA digestion. Its transforming activity[1] was plotted against digestion time. (A dependence was observed -- thus supporting this hypothesis.)

b.   Digest the fragments using BamH1 - which cuts in fewer places. Again a ten-fold reduction in transforming activity was observed. This further confirmed that the loss in transformational ability was due to the size of the pieces and not due to a property of $EcoR_I$.

F.   Repeat the digestion step -- limiting the process to partial digestion.

III.   Construct hybrid plasmids by mixing the partially $EcoR_I$-digested Phi-3-T DNA with $EcoR_I$-cut pSC101, followed by ligation with T4 ligase.

IV.   Transform E. coli with the plasmids. (See Chapter IV for the reasoning at this step.)

A.   Grow colonies and select for Tc^r. A control is done simultaneously with no plasmids added.

1.   Hypothesize that most of these colonies correspond to plasmids which have simply reclosed under ligase.

2.   Test for Phi-3-T DNA. Colony hybridizations indicates 8 percent of the Tc^r colonies have Phi-3-T DNA.

3.   Select for Thy^+. Two colonies are Thy^+.

B.   Starting again, grow colonies and select for Thy^+. Two colonies were obtained. Again, a control is done simultaneously with no plasmids added.

1.   Select for Tc^r. (Both colonies were Tc^r)

V.   Verify that the Thy^+ gene is harbored on a plasmid. Other plausible hypotheses should be tested. Possible hypotheses are: 1)Thy^+ is caused by reversion of Thy^- to Thy^+, or 2)the Phi-3-T thy gene has integrated into the E. coli chromosome.

A.   Remove the hybrid plasmids from the E. coli using a newly developed curing technique. The cured E. coli are distinguished by being Tc^s. All of the cured bacteria are also

Thy⁻. This tends to rule out the alternate hypotheses since it confirms that the Thy⁺ character is lost when the plasmid is lost.

B.   Propose further confirmation by reintroducing the plasmids into the cured bacteria. The same transformation frequency was observed in the cured bacteria as in the ones in which the plasmids had not been previously introduced.

C.   Propose further confirmation step of introducing plasmid into B. subtilis and testing for expression of Thy⁺. (Done below.)

VI.   Use heteroduplex analysis to examine the molecular structure of hybrid plasmids. (This time-consuming operation uses cRNA and EM.)

A.   Look for segments common to all of the transformed plasmids. Observe segment A is in all transformed plasmids. Observe unexplained hairpin in pFT33 in segment corresponding to segment A. This segment is longer and has distinct restriction sites.

B.   Hypothesize that segment A is from Phi-3-T. Confirm with heteroduplex mapping. Conclude that Thy is carried on segment A.

C.   Observe that segment A lies in two different orientations. Hypothesize that promoter control for Thy comes from Phi-3-T DNA and is part of segment A.

D.   Propose confirming experiment with Col El-amp plasmid. Choice of a different plasmid would disambiguate any special aspects of the original vector which might be involved.

VII. Transform B. subtilis with the hybrid plasmids.

A.   Explore interesting theoretical question: Does the topology of a plasmid (i.e. linear or circular) have any effect on the transforming activity of DNA?

B.   Test this question using BamH1 to linearize the plasmid. No difference in transforming activity of hybrid plasmids noted.

VIII. Extend the colony hybridization technique to B. subtilis. Involves showing: (1) that DNA can be detected when it is present and (2) (specificity) that it will not be detected when it is absent.

A.   Show (1) by hybridizing cRNA from Phi-3-T with B. subtilis strain lysogenized with the Phi-3-T.

8

B.   Show (2) by demonstrating absence of hybridization with pSC101 cRNA.

IX. Using the (now tested) _in situ_ hybridization procedure, perform the following measurements:

Test	Results Phi-3-T	pSC101
SB168	+	−
SB168 (with lysogenic Phi-3-T)	++	−
SB591 thy⁻	−	−
SB591 transformed with Phi-3-T	+	−
SB591 transformed with pFT23	+	−
SB591 transformed with pFT24	+	−
E. coli (with pSC101)	−	++

Conclude that these results further confirm the hybridization procedure. Also recognizing that SB591 is a mutated Thy⁻ derivative of SB168 (a standard strain of _B. subtilis_ with an introduced lysogenic Phi-3-T), conclude that the mutagenesis has deleted sequences of Phi-3-T.[1] Also note that _B. subtilis_ transformed with both pFT23 and pFT24 has kept the Phi-3-T DNA but not the pSC101 DNA. (Suggests an intriguing selection process at the molecular level.)

X.   Reverse the Southern hybridization procedure -- instead of electrophoresing the plasmids and checking for hybridization with Phi-3-T DNA, digest and electrophorese the Phi-3-T and check for hybridization with the pFT´s (using cRNA made from the pFT´s.) Unexpectedly, hybridization was observed in several bands. (Expected only one band corresponding to Thy.) Postulate repetition in the Phi-3-T genome. The experimenters attempted to confirm this result with heteroduplex analysis. This failed and the discrepancy was explained due to difficulties of this technique with large molecules.

----

[1]   (Other hypotheses have been offered since the papers were published.)

9

Chapter IV

Review of Knowledge Used in This Experiment

The following paragraphs step through the experiment described in Chapter III and emphasize the reasoning and the knowledge used to make the decisions and interpretations. This knowledge is highlighted by indentations in the text below. No attempt has been made to classify the knowledge here or to put it in a consistent form. The effort has been directed to writing down a first approximation to the knowledge with the intention of doing it more carefully at a later date after the scope of this knowledge is better understood. Thus, very high level strategy knowledge has been freely intermixed with very context specific knowledge in a potpourri of facts, directives, and rules of inference. We begin with the selection of the experiment.

IV.1   Proposing the Experiment

The knowledge used in proposing the basic experiment seems to be difficult to encapsulate. Part of the problem is that the considerations can be very broad -- involving political and regulatory considerations as well as the directions of long term research goals. Another problem is that many of the considerations seem to be fairly volatile -- what is a "hot" topic today will likely be less important tomorrow. In what follows, we will present some of the knowledge which seems to have been important in the proposal of this particular experiment along with the caveat that the appropriate place for MOLGEN's activity will undoubtedly be at a much lower level. No claim is made about the completeness of our characterization of the knowledge used at this level.

> Today, gene transfer is interesting --
> especially between species.

> Today, it is interesting to study whether gene
> control signals from one species are operative in
> another species.

These considerations derive from part of the long range research objectives of Lederberg's group. The particular experiment also requires a choice of species and genes. As was stated earlier,

attempts by this group to transfer a B. subtilis gene to E. coli had
been unsuccessful. Two members of the group attended a conference at
Cornell University in 1974 and brought back news of the bacteriophage
-- Phi-3-T. This phage was especially interesting because of its
ability to transform Thy$^-$ B. subtilis to prototrophy. Additionally, a
published reference was available which analyzed the EcoR$_I$ restriction
pattern of Phi-3-T[1]. It was suggested that DNA from this phage might
be used as the source of the Thy gene in the gene transfer experiment.
The following knowledge bears on that suggestion.

> It is easier to clone genes which can be
> obtained in high concentration and purity.

> Phage DNA can be obtained in high
> concentration and purity.[2]

A host species must also be chosen for a gene transfer
experiment. Much experience has been gained working with E. coli and
B. subtilis. Because so much is known about the genetics and
requirements of these organisms, they are among the organisms of choice
for many genetics experiments. Some particularly relevant facts
follow.

> When a species is available in strains with
> inactive genes[3], it may be a useful recipient for
> analogous genes from other species in a gene
> transfer experiment.

> There must be a means for incorporating the
> foreign DNA into the species so that it will be
> reproduced when the species grows.

> Plasmids and lysogenic viruses are typical
> vectors for introducing DNA between strains of
> bacteria.

> A large number of plasmids of E. coli have
> been characterized and are available.

----
[1] See [Wilson74].

[2] Phi-3-T DNA weighs 83 million, B. subtilis weighs 2.3 billion.
It is reasonable to get phage DNA in concentrations of 10^{11} pfu/ml
(Plaque forming units per milliliter.) As indicated already, Phi-3-T
was known to have a transferable Thy region.

[3] (that is, it is a well-characterized auxotroph)

The reverse experiment, namely transferring Thy (or some other marker) from E. coli to B. subtilis is also interesting. However, a cloning vehicle (eg. vector) must be found. This is currently an active area of interest.

IV.2   Establishing Some Subgoals

Most recombinant DNA experiments (of which this is an example) involve the following technical choices:

I. Choice of a Method for cutting DNA. (Alternatives follow.)

A. One of the restriction enzymes.

B. Physical shearing.

II. Choice of a Method for joining DNA segments.

A. A Ligase (Usually either E. coli ligase or T4-induced E. coli ligase.)

B. The (A:T) Terminal Transferase Method. This method involves putting polyA and polyT sequences on the ends of the DNA segments to be joined. Hydrogen bonds will form and hold the segments together when they are mixed in solution.

C. Molecular Adapters. This is a recent technique suggested at the last Miles conference. For situations where the DNA is cut by different restriction enzymes, the "sticky ends" of DNA will not match properly and segments cannot usually be joined. Molecular adapters are short segments of DNA with alternate sticky ends corresponding to two different restriction enzymes. They may be used to splice together fragments resulting from digestion by different restriction enzymes.

III. Choice of a vector for introducing the gene.

The following knowledge is relevant for these selections:

Restriction enzymes are useful for experiments where repeatable patterns need to be created. They can be used to create large numbers of identical DNA segments.

Shearing can be used for situations where all

known restriction enzymes inactivate the desired gene.

Shearing can be used to obtain a bank of adjacent genes.

T4 Ligase is capable of joining segments which are both flush-ended or segments with complementary ends ("sticky-ends").

T4 Ligase of reliable purity is currently available in Lederberg's laboratory.

E. coli ligase (not T4) can be used to join sticky ends, but not flush ended DNA.

A big advantage of the Terminal transferase method is that it insures that exactly one DNA region will be inserted.

Molecular adapters were not available. [4]

In the current experiment, the availability of T4 ligase and experience with restriction enzymes led to the establishment of the following subgoals for this experiment:

I. Isolate the gene on a segment of DNA of the right size having sticky-ends left by a suitable restriction enzyme.

II. Cut an appropriate plasmid with the same restriction enzyme so it has the same sticky ends.

III. Ligate the gene segments and plasmid segments to create the hybrid plasmids.

IV. Transform the E. coli with the hybrid plasmids.

In addition, the selection of a vector and restriction enzyme has to be made. The following knowledge appears to bear on the selection of the plasmid.

There should be good genetic markers for any genetic transfer method.

The   vector should   be small   and easily

----

[4] They now are available. See [Scheller77].

incorporated into a cell. (Many plasmids satisfy this requirement.)

It is a great advantage for purification steps if there are a large number of copies of the vector in each cell.

The vector should have a restriction site for one of the available restriction enzymes.

Punctuation signals must be available for the inserted gene. (These may be provided by the inserted segment or they can be near the restriction site of the vector.)

The utility of genetic markers on the vector derives from the following considerations:

Three kinds of labeling are typically used in experiments involving DNA manipulation -- radioactive labeling, density labeling, and biological labeling.

Biological labeling (genetic markers) offers the advantage of amplification via growth and selection in a medium.

In this experiment the Tc^r gene on pSC101 and pMB9 provide this means by way of the biological test for tetracycline resistance. The restriction enzyme should be picked with the following considerations:

If a later ligation step is planned, the restriction enzyme should cut leaving complementary ("sticky") ends.

If the restriction enzyme is known to cut (inactivate) a gene, it should probably not be chosen for use in an experiment to transfer that gene.

A restriction enzyme should be chosen which has a recognition site compatible with the size of segments desired (eg. four, five, or six base pairs).

IV.3   The First Part of the Experiment

As part of meeting the Subgoal I above, a sub-sub-goal is the extraction and purification of the Phi-3-T DNA. This should cause a selection among DNA-purification procedures. The method used in this experiment is a standard procedure. Some obvious but important rules apply here:

> Even if theory predicts something strongly, if a test is easy, do it. (You may be surprised.)

> Verify any important step (e.g. purification) with further confirmational tests.

> When there is a disagreement among measurements, or between a local measurement and a published measurement, find an alternate way to do the measurement.

> DNA purification steps should be checked for protein contaminants, RNA, and degree of intact molecules. Often it is appropriate to check the biological (e.g. transforming) properties too.

> An easy and fast test (OD 260/OD 280) is available for testing for the presence of protein in DNA.

> Electron microscopy can be used to test the degree of intact molecules.

The habit of checking and verifying steps is ubiquitous in experimental procedures. In this experiment, the UV test was performed although it is so standard that it is usually not reported. In addition, the degree of intactness and molecular weight were measured using EM.

Following the digestion with EcoR$_I$, it is possible to observe a restriction pattern and to measure the molecular weight of the Phi-3-T DNA using gel electrophoresis. In this case, a discrepancy was observed between the EM measurement of the molecular weight and the gel electrophoresis measurement. The following hypotheses were generated to explain the discrepancy.

I.   Incomplete gel resolution.

II. Loss of small DNA fragments which ran all the way through the gel.

III. Experimental error in measuring molecular weight.

IV. Repetition in the phage genome. (See Section IV.6.3.)

   The following knowledge was relevant in generating these hypotheses:

     Incomplete gel resolution leads to underestimation of molecular weight. (The usual assumption is that each band in the restriction pattern contains the same amount of material and corresponds to a single segment of the molecule. If resolution is incomplete, at least one band actually corresponds to more than one segment and the estimate of total molecular weight will be too low.)

     Loss of small fragments on the gel can lead to an underestimation of molecular weight. (This means that the electrophoresis has been run so long that the smaller fragments have migated all the way through the gel.)

     Accuracy of gel electrophoresis for measuring molecular weight in the linear part of its range is usually about two percent.

     For both electrophoresis and EM, a DNA standard must be run simultaneously to calibrate the measurement. In both cases it must be possible to unambiguously distinguish the standard from the molecules being measured. Different standards were suitable for these measurements.[5] There has been some disagreement about the value of the molecular weight for the standards.

   No further experimental effort was spent on discriminating between these hypotheses for the discrepancy. An alternate approach to measuring the molecular weight was proposed based on the following knowledge:

----

[5] SPP1 DNA cut with $EcoR_I$ for electrophoresis, and pSC101 (circular) for EM.

Incomplete gel resolution is often caused by having too many bands of DNA on the gel.

The number of bands in the gel is a function of both the restriction enzyme used and the DNA itself. (It depends on the location and number of restriction sites.)

BamH1 makes fewer cuts in Phi-3-T than $EcoR_I$. (Four instead of about thirty.)

However, the fragments of Phi-3-T DNA left after digestion with BamH1 are too large for accurate measurement with electrophoresis. Thus in this case, this idea does not provide a good alternate source for the molecular weight measurement.

As noted earlier, it is often worthwhile to check the transformational activity after a purification step and prior to a cloning step. In this case, the completely $EcoR_I$ digested Phi-3-T DNA was observed to transform B. subtilis, but at a 1000-fold decrease in efficiency compared to the uncut Phi-3-T DNA. The following hypotheses have been suggested to explain the loss of efficiency:

I. $EcoR_I$ cut the Thy gene -- thus damaging its transforming ability.

II.   Transforming activity decreases if the DNA segments are too small.

III. Edge effects.

The following knowledge is relevant to the generation of these hypotheses:

A gene which has been modified will usually function at an impaired efficiency.

Although the transformation process is not thoroughly understood, it is known that the process is influenced by DNA structural features. [6]

A recent thesis by one of Lederberg's students suggested that some genes function less effectively if the DNA is cut near the gene. [7]

----

[6] See [Notani74] for a review of what is known about this process.

[7] This has not yet been demonstrated conclusively.

The first hypothesis was disconfirmed using the following knowledge.

> If a gene can be shown to functional after its DNA has been completely digested by a restriction enzyme, the enzyme probably does not cut the gene.

> If no change is observed in the restriction pattern (in electrophoresis) for some DNA after a ten-fold increase in digestion time and enzyme concentration, the DNA [8]may be assumed to be digested to completion.

The hypothesis about fragment size suggests a course of action using the knowledge that:

> Sometimes experimental parameters can be changed to maximize efficiency.

> The completeness of digestion of DNA by an enzyme can be controlled by suboptimizing reaction rates. Many DNA segments after partial digestion of a restriction enzyme are larger than the segments left after complete digestion.

> Enzyme digestion conditions optimal for various purposes can be determined by studying the enzyme kinetics.

The experimenters decided to study the kinetics of transforming activity versus digestion. Conditions appropriate for a 10-fold reduction in transforming activity were found and used previous to the ligation step.

The hypothesis that the transforming activity depended on the size of the segments (and not on some property of $EcoR_I$) was further confirmed using the following knowledge:

> Different restriction enzymes will cut DNA into different numbers and sizes of pieces.

-----8--------
[8] Although this rule was cited in the papers, some exceptions to it are generally recognized. In the first place, there are inherent resolution limits which prevent some changes in restriction patterns from being observable in electrophoresis. Secondly, sometimes restriction sites can be covered by trace proteins.

This leads to the confirmation step using BamH1 as an alternate restriction enzyme. The assumption behind this step is that $EcoR_I$ and BamH1 will reduce the transforming activity by the same mechanism[9].

IV.4   A Brief Recapitulation

To recapitulate some of the logic here, we started out with an experiment for using Phi-3-T and E. coli based in part on some criteria for interestingness of the the experiment and appropriateness for the species involved. Criteria such as availability led to the selection of pSC101 and the restriction enzyme $EcoR_I$. The following subgoals were established:

I.   Isolate the Thy gene on a segment of DNA, that is, cut the Phi-3-T DNA so that the Thy gene will be located on a shorter segment of DNA. (Use a restriction enzyme which leaves sticky ends.)

II.   Cut the plasmid with the same restriction enzyme.

III. Ligate a mixture of the cut Phi-3-T DNA and some cut pSC101 plasmids resulting in some hybrid plasmids.

IV.   Transform the E. coli with the hybrid plasmids.

The first subgoal led us to a purification step and we paused to test the purity of this step. An unexpected discrepancy in molecular weight led us to some hypothesizing and checking. Then the necessity to test the transforming activity of the Phi-3-T fragments took us afield because the activity was unacceptably low. This led to the hypothesis that the loss of activity was related to the size of the DNA fragments. This hypothesis was tested and resulted in a modification to the plan -- partially instead of completely digesting the Phi-3-T DNA. (This modifies the second subgoal established previously.) We are now ready to pursue the last two subgoals.

IV.5   The Second Half of the Experiment

----

[9] It is possible that the two restriction enzymes could both reduce transforming activity of Phi-3-T DNA, but by different mechanisms such as exonuclease contaminants. The considerations and hypotheses that were generated and tested on this topic were not reported in the papers.

The third subgoal is to ligate the plasmids. We have already cited some knowledge necessary for the selection of T4 ligase. After ligation it is good practice to verify the success of the step. The following hypotheses could be tested:

I.   The Ligase sealed the DNA.

II.   Phi-3-T DNA has been incorporated in the plasmids.

III.   The Thy gene from Phi-3-T DNA has been incorporated in the plasmids.

The following knowledge is relevant to generating and testing these hypotheses:

> If the Ligase has sealed the DNA, it will form covalently closed circles.

> EM can easily and cheaply test whether circular loops of DNA are in the sample.

> Ligation theory [10] can be used to predict how many of the plasmids will incorporate extra DNA.

> Most of the plasmids will simply reclose without incorporating any Phi-3-T DNA.

> If circular DNA in seen in EM, this will constitute some evidence for successful ligation.

The following knowledge is relevant to the testing of the third hypothesis:

> Heteroduplex mapping could be used to confirm the incorporation of Phi-3-T DNA in the plasmids.

> Heteroduplex analysis is most applicable to a homogenous population of molecules. [11]

----

[10]   See [Dugaiczyk75].

[11]   The ligation products are a mixture of linear and reclosed pSC101 plasmids, Phi-3-T DNA fragments, and hybrid plasmids. The biological screening in the next step will help concentrate the molecules which include the Thy gene.

If the plasmids are used to transform the E. coli and amplified by growth (the next step in this experiment), there will be a large number of hybrid plasmids available for further testing.

These considerations led to using a simple EM test for successful ligation in the current experiment. Testing of the second and third hypotheses was deferred until after the transformation step.

The final subgoal in the basic experiment is to transform the E. coli with the hybrid plasmids. Like all important steps in an experiment, transformation needs to be checked. The following hypotheses about the results of the transformation could be differentiated:

I. No E. coli will have a $Thy^+ Tc^r$ phenotype. (All colonies will be $Thy^- Tc^s$, $Thy^- Tc^r$, or $Thy^+ Tc^s$).

II. Some E. coli will have $Thy^+ Tc^r$ phenotype.

A. These E. coli have no hybrid plasmids.

B. These E. coli have hybrid plasmids.

1. The phenotype is not conferred by the plasmids.

a. E. coli showing a Thy^+ phenotype are revertants.

b. E. coli showing a Tc^r phenotype are the products of contamination.

2. The Thy^+ character is conferred by the plasmids.

Testing of the first hypothesis illustrates the importance of biological markers on the plasmid -- Tc^r in this case.

Biological markers have associated tests for function deficiencies. [12] Plus phenotypes can be selected for directly.

If bacteria can grow in a medium lacking in an essential nutrient, they are synthesizing it for themselves.

If bacteria can grow in a medium with an

----

12 Minus phenotypes can be obtained using replica plating.

antibiotic present (e.g. Tc), they are resistant to
it at that concentration.

The Tc^s phenotypes could arise if none of the plasmids were
incorporated into the E. coli or if the gene for tetracycline
resistance has been inactivated. The Thy$^-$ phenotype would be expected
if either the Thymidine synthetase gene could not be activated in E.
coli or if it had been lost or damaged during the preparations. The Tc^s
Thy$^+$ phenotype would have been unexpected, but might have resulted if
either the tetracycline gene were damaged or the Thy gene had become
incorporated by some unanticipated mechanism. As discussed in Section
III.2, colonies with the Thy$^+$ Tc^r phenotype in fact were found.

The next hypothesis to be tested is whether the Thy$^+$ Tc^r
colonies contain hybrid plasmids. Three sources contribute evidence
relevant to this hypothesis -- a biological argument about phenotypes,
a colony hybridization step, and a heteroduplex analysis step.

The biological argument is somewhat indirect in that it
provides an opportunity for an easy counterexample. It is based on the
following knowledge:

> Genes which are close together will tend to
> stay close together through transformation,
> transduction, conjugation, etc. Genes which are
> far apart (eg. chromosomal vs plasmids) will stay
> far apart[13] and will generally not be co-
> transferred[13].

The fact that all of the E. coli which are Thy$^+$ are also Tc^r is
indicative that the Thy$^+$ phenotype is conferred by the plasmids.
However, the appearance[14] of any E. coli that were Thy$^+$ Tc^s would have
been quite unexpected[14] and a serious challenge to the claim that the
phenotype was conferred by hybrid plasmids. Genes linked in this
manner are especially useful in other experiments where there is no
direct way to select for the gene inserted on the plasmid.

The second source of evidence that there are hybrid plasmids is

---

[13] There are exceptions to this rule that are not relevant here.
For example, there is a sex-factor found in Hfr and F^+ E. coli strains.
These factors are conjugative (mobilizable) plasmids which can be
transferred to F$^-$ recipients. In this experiment, all of the E. coli
were F$^-$.)

[14] They would most likely be due to a reversion of Thy$^-$ to Thy$^+$.
This was disconfirmed by the control plate as discussed below.

from a colony hybridization step.   This technique involves using RNA
polymerase to make a template of radioactively labeled RNA which is
complementary to the   sequence we wish to detect. The bacteria are
grown on a filter, lysed (cell walls are burst), fixed, and then washed
and hybridized with the cRNA. The cRNA will bind to any complementary
DNA on the filter and can be detected autoradiographically. Again,
this evidence does not conclusively demonstrate that there are hybrid
plasmids -- only that DNA from Phi-3-T has been incorporated into the
Tc^r Thy^+ colonies. The DNA could conceivably be incorporated in some
other way.

The third source of evidence for hybrid plasmids is convin:ing
but time-consuming to perform -- heteroduplex mapping. This step
demonstrated the existence of hybrid plasmids and elucidated their
structure. (If either of the previous tests had failed, it might not
have been worthwhile to perform this test.)

The next hypothesis to be tested is whether the Thy^+ phenotype
is actually conferred by the plasmid. Some reason to test this
question is suggested by the following:

Genes   can   be   carried   on   the   bacterial
chromosome or on extra-chromosomal DNA (episomes)
such as plasmids or non-lysogenic phages.

Plasmids   are   occasionally   picked   up   as
contaminants from the air or medium.

The hypothesis that the Thy^+ character is the result of a
reversion is ruled out by a control plate where no plasmids have been
added (and no Thy^+ colonies appeared) along with the following
knowledge:

The Thy⁻ E. coli in this experiment have been
characterized as a deletion mutant.

Most deletion mutations very rarely revert.

Similarly, the hypothesis that the Tc^r phenotype was the result of
contamination was disconfirmed by an analogous control plate.

Disconfirming the specific hypotheses of other ways that the
phenotypes could have been achieved does not prove that the phenotype
is necessarily the direct result of the hybrid plasmid because we have
not ruled out all possible contrary hypotheses. However, the following
idea does provide a method for a fairly direct demonstration.

To show that A is the sole cause of B, Show
(1) that when A then B and (2) that when not A then
not B.

The pSC101 plasmids transform $\underline{E.}$ $\underline{coli}$ with
high efficiency.

The pSC101 plasmids may be removed from $\underline{E.}$
$\underline{coli}$ by a technique involving ethidium bromide,
tetracycline, and ampicillin.

Techniques   for   removing   or   transforming
bacteria with particular plasmids are expected to
continue to work after small segments have been
inserted into the plasmids.

These considerations led to the experimental step of removing
the hybrid plasmids by the technique above. $\underline{E.}$ $\underline{coli}$ without the
plasmid are identified by virtue of the fact that they are Tc^s. The
fact that the cells which were Tc^s were also Thy^- adds further credence
to the hypotheses above by establishing a linkage between the genes.
Finally, the plasmids removed from the Thy^+ Tc^r cells showed a high
transforming efficiency for both the $\underline{E.}$ $\underline{coli}$ which never had the hybrid
plasmids, and for $\underline{E.}$ $\underline{coli}$ from which they had been removed. The latter
colonies were in every tested respect identical to the former. In
particular, when the latter cells were mixed with the hybrid plasmids,
they were transformed to Thy^+ Tc^r at the same high efficiency as the
original cells.

IV.6   Some Experimental Fishing Trips

At this point the experiment could very well have been
terminated. A great deal of evidence had confirmed the successful
transferring and expression of a gene from Phi-3-T to $\underline{E.}$ $\underline{coli}$. Thus,
although some of the next steps in the research were partially
motivated by a search for further confirming evidence that the Thy^+
gene was on the plasmid, the opportunity was taken to perform some
simple related experiments.

IV.6.1   Transforming B. subtilis with the Hybrid Plasmids

Although the experiment as described above has already

presented strong evidence for the successful transfer of a gene to E. coli, the following knowledge may have suggested some further steps.

A gene transferred to a new host can be tested for modifications by reintroduction into its donor.

If a particular gene is known to function in a particular species of bacteria, attempts to clone that gene into a deficient strain will probably succeed without complications involving gene expression[15].

The Thy$^+$ gene is carried by the bacteriophage Phi-3-T and expressed in its host B. subtilis when the phage is incorporated lysogenically.

B. subtilis often incorporates DNA which it encounters in its environment (eg. it is highly transformable).

It would be interesting to know more about the mechanism which B. subtilis uses to incorporate and control foreign DNA.

This suggests that it would be interesting to try to transform Thy$^-$ B. subtilis with the hybrid plasmids (termed pFTs). Successful transformation would also provide further confirmation that the Thy$^+$ character of the transformed E. coli was plasmid borne. The experiment was performed and it was observed the transformational activity was 10^{-6}. At this point, the following knowledge appears to have been active.

(It is interesting to know what factors contribute to the optimization of important processes.)

The incorporation of DNA by bacteria is a process which is not well understood. It is not known what structural features may influence this process.

EcoR$_I$ cut Phi-3-T DNA is linear.

----
[15] There are sometimes complications involving dosage and locations of promoter sites.

The pFT's are circular and contain both pSC101 DNA and Phi-3-T DNA.

The transforming activity of the Phi-3-T was 10^{-5} and the transforming activity of the circular or BamH1 cut pFT's was 10^{-6}.

If the pFT's are cut with $EcoR_I$ the pSC101 DNA will become disconnected from the Phi-3-T DNA (because the plasmids were ligated at $EcoR_I$ sites.)

This suggests that it would be interesting to decide which factors (differing between the pFT's and $EcoR_I$ cut Phi-3-T DNA) determine the reduced transforming activity. The considerations above led to the experiment of cutting the pFT's with BamH1. No change in transforming activity was noted -- although the cutting linearizes the plasmids. It may be hypothesized that the distinction between linear and circular DNA segments makes no difference to the B. subtilis in this case.

IV.6.2   Extending the Colony Hybridization Technique to B. subtilis

Rapid and reliable assays for important properties are worth developing.

In situ hybridization is a good assay for incorporated DNA and is an important assay for cloning experiments.

In situ hybridization has only been tested for DNA incorporated by E. coli.

Sometimes techniques which are useful for one species can be extended without much effort to another species.

These considerations suggest that the in situ hybridization technique could be extended to work for B. subtilis as well as E. coli. Some further experiments related to the step in the previous section would then become easy to do.

To validate an assay, it is necessary to demonstrate both a capability for detection and specificity[16].

When a phage is lysogenic with a bacterial host, the phage DNA is incorporated into the the bacterial chromosome.

SB168 is a standard and available strain of B. subtilis which can be lysogenized with Phi-3-T.

This knowledge suggests that the *in situ* hybridization procedure may be tested if Phi-3-T lysogenized SB168 shows positive hybridization technique with Phi-3-T and negative hybridization with pSC101.

When a new technique is validated, it is interesting to try a variety of test cases. (They may suggest further research.)

SB591 is a Thy⁻ derivative of SB168 available in Lederberg's laboratory.

In fact, four out of seven of the test cases offered surprises. The first case was SB168 tested (as was each test case) with Phi-3-T-derived cRNA and pSC101-derived cRNA. As predicted, no homology with pSC101 was observed but the Phi-3-T homology was a surprise (not yet explained.) The second test case was the same as the validation test for the method and offered no surprises. Next the mutated SB591 was tested. Interestingly, it showed no homologies with either cRNA which suggests that the mutation has deleted regions homologous to Phi-3-T. The next three test cases involved transformations of SB591 to Thy⁺ -- by two pFT's and by Phi-3-T DNA. As expected, homologies (presumably due to the Thy gene) were detected in each case. Surprisingly, no homology was seen with the cRNA from pSC101 -- suggesting that the B. subtilis has deleted segments of DNA arising from the pSC101 component of the plasmid. (Further research will be required to explicate this.) Finally the test case of E. coli with pSC101 plasmids offered no surprises.

To summarize some of the knowledge implied by these test cases and conclusions:

----
[16] This is a special case of the earlier rule. To show that A is the cause of B (i.e. that B measures A), show (1) that when A then B and (2) that when not A then not B.

When a bacterial strain is transformed by a segment of DNA and later tests show homologies for only some parts of the introduced DNA, this is evidence for a selective process.

A mutation may involve structural rearrangements, such as inversions, insertions, substitutions, or deletions of segments of DNA.

When a homology test, which is done for a strain and a mutated version of that strain, shows a loss of homology to some test DNA, this constitutes evidence for deletion of DNA by the mutagenesis process.

A confirming test for a deletion mutant is absence of reversion under selective pressure.

IV.6.3   Back Hybridizing the pFT's to Phi-3-T DNA

The Southern method[17] is useful for finding regions of homologies between two samples of DNA.

Homology tests between DNA samples A and B can be performed by testing A against regions of B, or by testing B against regions of A.

In a previous step, the pFT's were tested against Phi-3-T DNA to provide evidence for insertion. Since it is just as easy to reverse the process, it was done and the surprising result was that pFT cRNA's hybridized with several Phi-3-T DNA bands. The following knowledge was used in making some tentative conclusions:

When a segment of cRNA (or cDNA) shows homology to several different parts of a DNA molecule, this constitutes evidence for repetition in the molecule.

When a DNA molecule is completely digested by a restriction enzyme, different bands in the restriction pattern correspond to distinct regions of the molecule.

----
[17] See [Southern75].

Heteroduplex mapping is useful for detecting
molecular rearrangements in DNA segments.

An alternative method to analyze repetitions -- heteroduplex
analysis -- revealed only one region of homology between the pFT and
Phi-3-T. Furthermore, the hairpin region in pFT33 (which could arise
from an inserted and inverted repeated sequence) remains unexplained.
These discrepancies have yet to be explained.

<u>Chapter V</u>

<u>Some Thoughts About the Knowledge Used in This Experiment</u>

Chapter IV stepped through the experiment and highlighted the knowledge that was active at critical decision points. In this chapter, a few general observations will be made about the knowledge and its use in planning for this experiment.

V.1   <u>General Observations</u>

<u>There is a large body of diverse knowledge used in this experiment.</u>

Approximately one hundred entries in the form of relevant facts or guidelines were highlighted in Chapter IV. When we actually try to fill out this knowledge and formalize it as specific entries in a MOLGEN knowledge base, it will undoubtedly swell considerably. The diversity of this knowledge suggests that a great deal of research effort in MOLGEN will be devoted to questions of representation.

<u>Much of the planning appears to be event driven.</u>

The experiments described here reflect a combination of goal driven behavior and event driven behavior. (The word "experiment" itself suggests that the procedure is somewhat <u>tentative</u> and intended to elucidate an unknown effect or law.) If there were no goals, behavior might seem very erratic and follow no general course. If there is no event driven component to the planning process, then the experimental procedure must admit no feedback or change of plans as a result of the observations. Thus, no advantage will be made of fortunate observations. What is being suggested here is that the planning in this experiment involved far more exploitation of events and changes of plan according to events than the authors had anticipated. The importance of a combination of goal driven and event driven processes in problem solving has been discussed in the artificial intelligence literature[1].

---
[1]   See, for example, [Erman76] or [Engelmore77].

One of the significant events in this experiment resulted from the observation of the low transforming efficiency of the $EcoR_I$ cut DNA on B. subtilis. This led to some hypotheses about the optimum size of DNA fragments for transformation, some confirmational measurements, and finally to an alteration of the high level plan. (The DNA was only partially digested instead of being completely digested before ligation with the plasmids.) Finally, some events did not lead to changes in the experiment but did contribute to the wealth of conclusions which could be drawn. For example, the observation that inserted segments in different hybrid plasmids were oriented in opposite directions led to a conclusion about promoter control. Similarly, the observation that pSC101 DNA was not incorporated into B. subtilis led to a tentative hypothesis about a selection process and suggests an area for further research. Unlike the observation about segment orientation, this observation was not anticipated.

One explanation for some of the event-driven character of the experiment is the fact that planning must take place even though the knowledge is incomplete. For example, when it was proposed that an alternate measurement of molecular weight could be obtained by using BamH1 instead of $EcoR_I$ to digest the Phi-3-T DNA before electrophoresis, the number of restriction sites and size of the Phi-3-T fragments were not known in advance. Thus planning steps must be proposed tentatively and checked after completion. The incompleteness of the knowledge may take the form of unknown properties of the laboratory techniques as well as unknown attributes of DNA structures. For example in a recent laboratory meeting, the question was asked whether a particular enzyme used in an experiment was precessive (i.e. whether it tends to remain attached to a single molecule). If in fact the enzyme were precessive, a failure in the experiment could be explained and the experimental procedure could be appropriately altered.

The geneticist is opportunistic and tries to make discoveries.

Thus, not only is the planning process largely event driven but sometimes steps are taken somewhat outside the plan of the experiment to make a possibly interesting observation. This kind of behavior reflects the convenience of making certain interesting observations while the equipment is set up. Often this is done to verify the successful completion of an experimental step, but sometimes the observations seem to correspond more to fishing for interesting possibilities. One example was the linearization of the hybrid plasmids with BamH1 to see if topology was important in the incorporation of the plasmids in B. subtilis. Another example was the transformation of B. subtilis by the hybrid plasmids. (Better evidence

was already available that the Thy gene was plasmid borne and that the transformation of E. coli had been achieved.) Several examples of this search for possibly interesting observations were presented in Section IV.6.

Hypothesis formation is an important activity in planning experiments.

Hypothesis formation is especially critical when an experimental prediction fails -- that is -- when the unexpected is observed in an experiment. For example, a difference in homology led to an hypothesis about the mutagenesis of SB591. In this case, hypothesis formation could be described by a single rule of evidence. Other cases of hypothesis formation are more involved. In such cases, the knowledge of the limitations and effects of laboratory techniques is likely to play a role in the formation of hypotheses. One example of this has already been cited -- the hypothesis that the low transformational activity of completely $EcoR_I$ digested Phi-3-T DNA was caused by the small size of the remaining DNA segments. Similarly, when the discrepancy in molecular weight was discovered early in the experiment, hypotheses had to be generated which could explain the differences among different sources for the measurement.

In many of these situations, the generation of the hypotheses can be understood as a systematic checking of the assumptions used in the model to make the disconfirmed prediction. Systematic generation of plausible hypotheses is clearly one of the most important processes in the experimental science. Some effort in this direction could prove interesting for MOLGEN.

It is often difficult to determine when there is enough evidence.

From one point of view, this observation is equivalent to the previous one. There is sufficient evidence when all competing hypotheses have been ruled out. Practically, there is no way to be certain that all plausible hypotheses have been considered, and there are many hypotheses which are too farfetched to merit serious examination. The problem arises from the fact that often no laboratory technique is available to directly measure the item of interest. Measurements have to be interpreted and techniques are subject to occasional failure. This generally leads to a fairly conservative approach to experimental proof so that several confirming measurements are made when the result is important.

Perhaps the best example of rather careful confirmation and differentiation between hypotheses is the reasoning involved in verifying the success of transforming E. coli. It was deemed insufficient to show merely that the E. coli were Thy$^+$. Rather the evidence was sifted to see precisely whether it indicated (1) that some E. coli were Thy$^+$ and (2) that the plasmid contained an insert of Phi-3-T DNA and (3) that the Thy$^+$ character was conferred by the plasmid. This careful testing reflects an ability to differentiate between hypotheses which could explain some of the tests individually without necessarily confirming the central hypothesis.

Thus, a great deal of evidence was gathered in support of the most important conclusion -- the successful transfer of the Thy gene to E. coli. Comparatively less effort was spent in checking the validity of the colony hybridization technique for B. subtilis, even though several surprises were found when it was used investigatively. Interestingly, a failure in the established heteroduplex analysis technique was postulated when it failed to confirm a prediction about repetition in the phage genome. This suggests (1) that even established techniques are occasionally suspect and (2) that the most important conclusions merit the most redundant checking.

V.2   Some Proposed Important Parameters

In reviewing the knowledge expressed in Chapter IV, it appears that some concepts have broad application.

Importance	Determines how worthwhile it is to pursue certain objectives at this time (like AM's^2 "interestingness").
Effort	How expensive or time-consuming it is to pursue certain objectives. (Resource limitations.)
Certainty	Determines believability of inferences and observations (like MYCIN's ^{3} certainty factors).

----
^{2} See [Lenat76].

^{3} See [Shortliffe76].

Safety   Is the experiment safe and does it fit
                    within regulatory guidelines?

Availability   Availability of materials is a major
                    consideration.

        This list is by no means complete and several specializations
may prove useful. For example, one can distinguish between items which
are important in a particular context, and those which are important
generally.   Something can be   important in the sense   of being
fundamental or in the sense of being novel or suprising. Planning
decisions may make considerations of more domain dependent notions --
for example yield, purity, or shelf life.

        Many of these parameters appear in a variety of decision making
contexts. For example, importance and effort are useful in deciding on
very long range objectives as well as short term and event driven
decisions. Important conclusions merit the most careful (redundant)
verification.   Unpredicted observations must be examined for their
importance (interestingness, novelty, and impact on genetic theory)
before effort is spent to explain them or perhaps even base further
research on them.

        Many combinations of these parameters are especially relevant
in decision making. For example, it is probable that resources should
be allocated to an experiment which has a high importance and a low
effort. An example of this is the linearization of the hybrid plasmids
in this experiment. If the verification of a hypothesis has a high
effort while a low effort counterexample can be tested, the
counterexample may be tested first even if it is unlikely. This is the
source of many of the "controls" which are routinely run.   In cases
where both the importance and the estimated effort are high, resources
are less likely to be allocated . An example of this would be a study
into the intriguing mechanism by which B. subtilis used some form of
molecular selection to reject the pSC101 part of the hybrid plasmid but
accept the Phi-3-T part of the plasmid. (This study could be a
genetics thesis itself.) When hypotheses are being verified, an
important hypothesis will merit considerable verification even if its
plausibility is already quite high. A less important hypothesis of
equal certainty will get somewhat weaker confirmation.   An unimportant
hypothesis of high certainty will probably get little attention. Other
interesting combinations of parameters include such things as important
and easy but unsafe experiments.   Much needs to be learned about when
it is optimal for considerations of these different parameters to enter
into the planning process. It seems that the characterization of these
parameters for decision making could be an interesting and significant
part of MOLGEN research.

        A cost/benefit analysis is required in general.

V.3   Rule Classifications

As previously mentioned, this paper made no attempt to categorize the rules used in planning the experiment. Rather, the rationale behind each experimental step was described by invoking rules at a variety of levels of detail. For example, in Section IV.6.2, a rule specifying the importance of verifying an important result is immediately followed by another rule which details the presence of a particular bacterial strain in the lab stock collection. This jumping from extremely high-level rules of scientific decision making to very specific statements of available resources needs to be examined in much greater detail.

It appears to be of critical importance to systematize the rules used in planning experiments. We have summarized a subset of this domain in describing the cloning experiment. However, many other possible strategies exist and many other experimental goals are possible. A proposed method of gaining knowledge about alternatives is by examining other experiments in detail. For this, a coherent structure needs to exist in order to delineate what rules, and at what levels of detail, are needed. To this end, a more formal notation is highly desirable. Our next step would seem to be generating a classification scheme for the rules used in this experiment with a view towards expanding the structure to including rules from other published experiments.

V.4   Concluding Remarks

A surprise in working on this set of experiments was the extent to which considerations outside the usual set of hierarchical planning ideas entered into the experimental planning. Far more of the decision making was event driven than had been anticipated and it remains to be seen whether this situation is characteristic of the current molecular genetics domain. In addition, some basic scientific activities -- such as hypothesis formation and testing, which had not previously received much attention in MOLGEN -- now appear to be quite important.

One of the important benefits of this exercise has been to make explicit some of the domain knowledge. One clear lesson from this exercise has been a greater realization of the importance of including some knowledge about biological function. Finally, more work needs to be done to categorize and represent the knowledge described in this report and to determine compact subsets of genetic knowledge which will provide the richest material for artificial intelligence research.

Appendix I

Bibliography

[Dugaiczyk75] Dugaiczyk A., Boyer H.W., Goodman H.M., Ligation of EcoR$_I$ Endonuclease-generated DNA Fragments into Linear and Circular Structures, Journal of Molecular Biology, 96, pp 171-184 (1975)

[Ehrlich76] S.D. Ehrlich, H. Bursztyn-Pettegrew, I. Stroynowski, and J. Lederberg, Expression of the thymidylate synthetase gene of the Bacillus subtilis bacteriophage Phi-3-T in Escherichia coli, Proceedings of the National Academy of Sciences USA, 73:11, pp. 4145-4149, (November 1976)

[Ehrlich77] S.D. Ehrlich, H. Bursztyn-Pettegrew, I. Stroynowski, and J. Lederberg, Cloning of the Thymidylate Synthetase Gene of the Phage Phi-3-T, Proceedings of the X Miles Conference, in press.

[Engelmore77] Engelmore R.S., Nii H.P., A Knowledge-Base System for the Interpretation of Protein X-Ray Crystallographic Data, Computer Science Department Report No. Stan-CS-77-589, Stanford University

[Erman76] Erman L.D., Overview of the HEARSAY Speech Understanding Research, Working Papers in Speech Recognition -IV- the HEARSAY II System, Carnegie-Mellon University, Computer Science Speech Group, (1976)

[Lenat76] Lenat D.B., AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search, PhD Thesis Computer Science Department, Stanford University (1976)

[Martin77] Martin N., Friedland P., King J., Stefik M.J., Knowledge Base Management for Experiment Planning in Molecular Genetics, Heuristic Programming Project HPP-77-19, Computer Science Department, Stanford University (1977) (Submitted to the Fifth International Conference on Artificial Intelligence.)

[Notani74] Notani N.K., Setlow J.K., Mechanism of Bacterial Transformation and Transfections, in Cohn W.E. (ed.) Progress in Nucleic Acid Research in Molecular Biology, 14 (1974)

[Platt64] Platt J.R., Strong Inference, Science 146, pp 347 (1964)

[Scheller77] Scheller R.H., et. al., Chemical Synthesis of Restriction Enzyme Recognition Sites Useful for Cloning, Science 196:177 (1977)

[Shortliffe76]   Shortliffe   E.,   MYCIN: Computer-based   Medical
    Consultations, New York: American Elsevier (1976)

[Southern75] Southern E.M., Detection of Specific Sequences Among DNA
    Fragments Separated by Gel Electrophoresis, Journal of Molecular
    Biology 98 pp 503-517 (1975)

[Stefik77] Stefik M.J., Martin N., A Review of Knowledge Based Problem
    Solving as a Basis for a Genetics Experiment Designing System, CS
    Report 77-596, Computer Science Department, Stanford University
    (1977)

[Wilson74] Wilson, G.A., Williams, M.T., Barney, H.W., and F.E. Young,
    Characterization of Temperate Bacteriophages of Bacillus Subtilis
    by the Restriction Endonuclease EcoR_I: Three Different Temperate
    Bacteriophages, J. Virol 14:1013 (1974)